An Adaptive Content Distribution Protocol for Clustered Mobile Peer-to-Peer Networks

Long Vu, Rahul Malik, Qiyan Wang, Klara Nahrstedt
Department of Computer Science, University of Illinois
201 N. Goodwin Avenue, Urbana, IL61801, USA
Email:{longvu2,rmalik4,qwang26,klara}@illinois.edu

***Abstract*—In this paper, we show that the clustered mobile peer-to-peer (P2P) networks exist in numerous scenarios where mobile users collaborate to improve content distribution services. In order to understand the clustering behavior of nodes in clustered mobile P2P networks, we present a probabilistic model for path selection and use the Mobius Tool to study the clustering behavior of different movement areas. Our study shows that the cluster size distribution follows an exponential function. We then design an adaptive protocol, which blends cellular and P2P (i.e., wifi or Bluetooth) communications of the mobile devices and leverages the exponential-cluster-size function to improve content distribution services. With our protocol, mobile nodes periodically sample the current cluster size and predict the future cluster size using the exponential function. Then, nodes apply the predicted cluster size function to calculate the available data in P2P channel using Online Codes and tune the cellular download timer adaptively to meet the file download deadline. The simulation results show that our adaptive protocol achieves much higher performance than the non-adaptive protocol by reducing the downloading load on the cellular channel by 4% $\sim$ 10%, and significantly reducing message overhead. Simulation results also confirm that our protocol adapts well to network dynamics since when the nodes get closer to the destination, the cluster size function is predicted more accurately.**

I. Introduction

The proliferation of mobile devices and swift advances of wireless technology lead to the emergence of pervasive computing paradigms. Modern mobile devices today, besides their traditional low-bandwidth, long-range cellular communication, already come equipped with additional wireless interfaces such as IEEE 802.11 (wifi) and Bluetooth. These interfaces can be leveraged for high-bandwidth, short-range (i.e., ad hoc/peer-to-peer) communication without any additional investment in the cellular infrastructure. Therefore, the peer-to-peer (P2P) communication becomes promising to exchange data among mobile devices and reduce the load on the cellular channel.

Some recent research projects have focused on combining cellular and P2P connectivity on cellular devices [1], [2], [3]. These approaches are based on the construction of a tree of ad hoc nodes rooted at high-data-rate proxy nodes connected to the cellular network. The maintenance of an ad hoc tree structure incurs a high communication overhead under network dynamics. Other projects [4], [5], [6] apply Bittorrent-like protocols to exchange data when nodes are in contact. However, these protocols are not adaptive to network dynamics. Moreover, performance of these protocols depends significantly on the movement pattern of mobile nodes, which depends on the movement of mobile users. Meanwhile, current mobility models [7], [8], [9] remain unrealistic since they do not capture the socioeconomic characteristics in human movement [10], [11].

Recently, a large-scale measurement study on human movement in Koblenz city [12] reveals two findings: (1) people do not always take the shortest path to the destination, instead people may choose different paths to the same destination although they start the trip at the same location, and (2) in a city, the movement of people shows clustering behavior, that is, the cluster centers have a higher density of people than other areas. In this paper, we explore how to use these human movement findings to greatly improve content distribution services among mobile users.

Particularly, we first show that the clustering behavior as indicated in [12] exists naturally in numerous scenarios where the mobile devices carried by participants form the *clustered mobile peer-to-peer networks*. We then study the clustering behavior of the clustered mobile P2P networks by proposing a probabilistic model for path selection, which assigns a higher probability for a shorter path to the destination. Relying on the probabilistic model, we use the Mobius Tool [13] to study the node distribution at the steady state of different clustered mobile P2P networks. Our study shows that the cluster size distribution follows an exponential function.

Then, based on the exponential-cluster-size function, we present a novel adaptive protocol that blends different wireless interfaces of the mobile devices to improve content distribution in clustered mobile P2P networks. With our protocol, mobile users download content from a content server via the cellular channel and at the same time exchange downloaded data via the P2P channel. When moving toward their destination, mobile nodes periodically obtain the current cluster size and predict the future cluster size using the exponential function. Then, mobile nodes apply the predicted exponential-cluster-size function to calculate the available data in P2P channel using Online Codes [14] (erasure coding) and tune the cellular download timer adaptively to meet the file download deadline. We evaluate our protocol using simulation and the results show that our adaptive protocol achieves much higher performance than the non-adaptive protocol. Particularly, our protocol reduces the downloading load on the cellular channel by 4% $\sim$ 10%, meets the file download deadline, and significantly reduces message overhead. Simulation results also confirm that our protocol adapts well to network dynamics since when nodes get closer to the destination, the cluster size function is predicted more accurately.

The paper is organized as follows. We first present the clustered mobile P2P networks in Section II. Then, we present our probabilistic model for path selection of nodes and study the steady state of clustered mobile P2P networks using the Mobius Tool in Section III. From the modeling process, we find that the cluster size distribution follows an exponential function, which motivates us in designing an adaptive content distribution protocol using Online Codes and the exponential-cluster-size function in Section IV and Section V. Then, we present the simulation results and related work in Section VI and Section VII, respectively. Finally, we conclude the paper in Section VIII.

II. Clustered Mobile Peer-to-Peer Networks

Data projected from the measurement study of pedestrian movement in Koblenz city [12] shows that the movement of people creates clustered areas. Particularly, in Koblenz city, the shopping street areas, Point of Interest areas, and attraction areas have a much higher density of people than other areas. In the context of mobile networking, the network formed by people's cell phones in such clustered areas is called the *clustered mobile P2P network*. In this network, the node distribution is not uniformed: clustered areas have much higher density of nodes than other areas and the density decreases as nodes are farther away from the cluster centers. We find that the clustered mobile P2P networks indeed exist naturally in numerous scenarios.

The first class of scenarios can be found in the context of location-dependent events. For example, pedestrians are moving toward the location of the same social event such as an outdoor concert, a football match, and a costume festival. When walking toward the event location, people may use cell phones to download a video file about the event content from a server via the cellular channel. Closer to the event location, there are more and more people heading toward the event location and downloading the event content. In this case, a clustered mobile P2P network can be formed by these cell phones that exchange downloaded data via the P2P channel to speed up the download process. Similar scenario exists when drivers drive toward the event location and want to access the event content using their cell phones. In this case, the connectivity among nodes is more dynamic since cars move faster than pedestrians. However, the overall phenomenon is similar since closer to event location there are more cell phones downloading the event content.

The second class of scenarios exists when pedestrians or cars move toward the same area and download the location-independent content. Previous studies [15], [16] show that in city workers move from residential areas to the office areas in the morning and move in the opposite direction in the afternoon. This renders an opportunity for collaborative downloads. For example, when commuters drive their cars from homes to the office area in the morning, they may use cell phones to download video files such as the breaking news, local weather information, and local commercial advertisement. Closer to a particular office area, there are more workers heading to this area and their cell phones can form a clustered mobile P2P network to download the video file from the server via the cellular channel, and exchange downloaded data via the P2P channel. Again, when commuters are driving back homes from their offices in the afternoon, they can form another clustered mobile P2P network to accelerate their downloads. Similar phenomenon exists when pedestrians are walking toward the same area of the city such as the city center and shopping streets. In this case, pedestrians may collaboratively use cell phones to download the video file of local news or the city attraction guide. Again, a clustered mobile P2P network can be formed by these cell phones to improve the downloads.

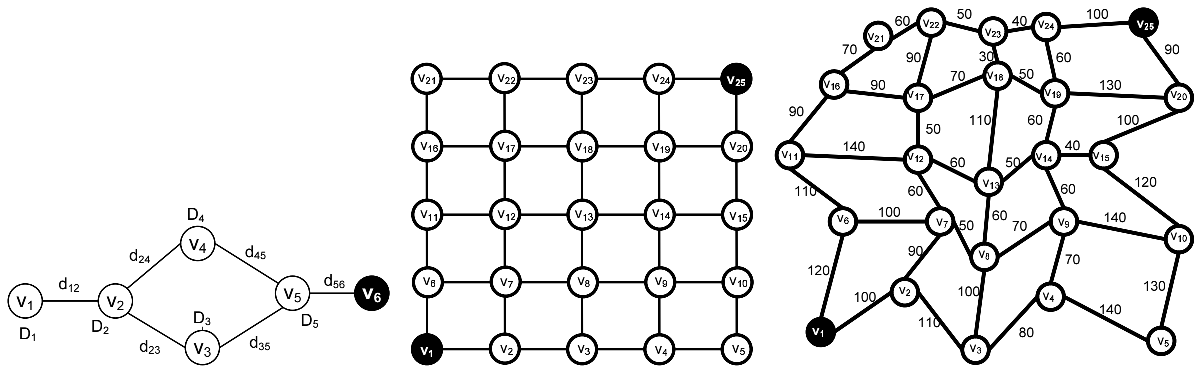
In the above scenarios, the clustering behavior exists due to the common targeted destination and common content interest. This clustering behavior differs from the Clustered Mobility Model in [17] since the authors in [17] state that there is only one node traveling between two adjacent clusters. In our clustered mobile P2P networks, there might be multiple nodes traveling among cluster centers and the density of nodes decreases when the distance to the cluster centers increases. We believe that this clustering behavior naturally occurs in reality, and offers a novel opportunity to improve the performance of content distribution services in clustered mobile P2P networks. In this paper, we first study the clustering behavior in clustered mobile P2P networks and then we exploit the result of the study to design an adaptive content distribution protocol.

III. Modeling the Clustering Behavior

In this section, we first model the path selection of an individual mobile node. Then, we study the clustering behavior resulted from the movement of multiple mobile nodes in different movement areas.

A. Probabilistic Model of Path Selection

We consider a mobile node n moving within a physical movement area, which consists of multiple street segments and intersections. Each street segment has two intersections at two ends. To simplify the model, we assume that the



(a) Sample Street Configuration (b) Manhattan Street Configuration (c) Random Street Configuration

Figure 1. The graph representation of different movement areas with different street configurations

destination is also at an intersection. From its current location, node n follows consecutive street segments toward the destination. We use an undirected graph $G =< V, E >$ to present the movement area as follows. The i^{th} intersection in the movement area is represented by a vertex $v_i \in V$. Meanwhile, a street segment connecting the i^{th} and j^{th} intersections is represented by an edge $(v_i, v_j) \in E$. An edge $(v_i, v_j) \in E$ has a weight d_{ij} representing the physical distance between v_i and v_j, hence we have $d_{ij} = d_{ji}$. Further, each vertex $v_i \in V$ has a value D_i representing the shortest distance from v_i to the destination. Figure 1(a) shows an example of the graph representation of a movement area. In this figure, v_6 is the destination and mobile nodes from other vertices will move toward v_6, and each vertex maintains the shortest distance to v_6.

As shown in [12], people do not always choose the (same) shortest path to the destination although they may start their trips at the same location, since people with different socioeconomic characteristics may prefer different paths [18]. To represent this behavior, we present a probabilistic model, which assigns a higher probability to the shorter path to the destination so that nodes can choose different paths to the same destination. We believe our probabilistic model represents the path selection behavior of humans better than previous mobile models [7], [8], [9] since it introduces the flexibility into the path selection process.

In our model, a path toward the destination consists of vertices in V. We define the transition probability p_{ij} as the probability that the node n selects v_j as the next vertex in the path toward the destination if n currently stays at vertex v_i. The next step is to calculate p_{ij}. Let N_i be the set of adjacent vertices of v_i, and each vertex of N_i has a shorter distance to the destination than v_i. Formally, $N_i = \{v_j : D_j < D_i \ AND \ (v_i, v_j) \in E\}$, and we have:

$$p_{ij} = \begin{cases} \frac{\frac{1}{d_{ij}+D_j}}{\sum_{v_k \in N_i} \frac{1}{(d_{ik}+D_k)}} & , \forall v_j \in N_i \\ 0 & , \forall v_j \notin N_i \end{cases} \quad (1)$$

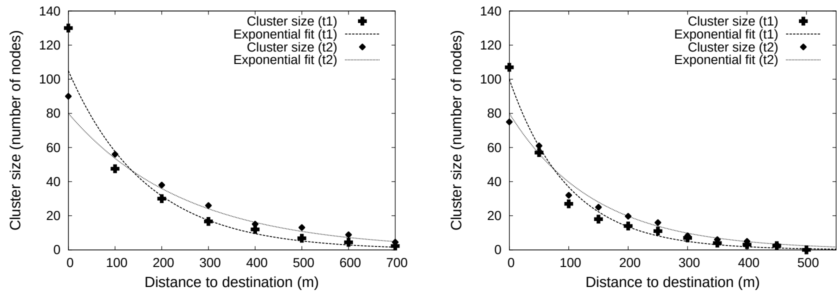
In Figure 1(a), if node n currently stays at v_2 and its destination is v_6, then n selects v_3 as the next vertex with the probability $p_{23} = \frac{\frac{1}{d_{23}+D_3}}{\frac{1}{d_{23}+D_3}+\frac{1}{d_{24}+D_4}}$.

B. Clustering Behavior

In this section, we study the clustering behavior of two different movement areas: Manhattan street configuration and Random street configuration as shown in Figure 1(b) and Figure 1(c). We are interested in clustering behavior at the steady state of the system since we believe the steady state provides insightful movement patterns of the system. In order to study the steady state, we specify two destinations for each movement area. To simulate the movement of nodes and study the clustering behavior, we use Mobius Tool [13], which is the Stochastic Activity Network and has been used widely for modeling the behavior of computer, distributed, and networking systems.

For the Manhattan street configuration, the distance between two adjacent vertices is $l = 100(m)$. Meanwhile, for the Random street configuration, we specify the distance with the edge of the graph. For each movement area, we have 500 mobile nodes moving to the destinations v_1 and v_{25} with a random speed s in the rage [1.0, 2.0] (m/s). For each vertex in G, we create one atomic model in Mobius and specify the transition probability of the case activity in this atomic model as the probability calculated by the Equation (1). The duration a node n stays in the activity of the atomic model is $\lceil l/s \rceil$ seconds, which is the traveling time node n spends to travel the corresponding street segment in the movement area. On arriving at one destination, nodes stay for a random staying period and then move to the other destination. For each movement area, we study the impact of the staying period at the destinations on the cluster size distribution and we consider two cases: the first case has staying period t_1 and the second case has staying period t_2 where $t_1 > t_2$.

Figure 2 shows the cluster size distribution in two movement areas with $t_1 = 4t_2$. From this figure, we see that the



(a) Manhattan Street Configuration

(b) Random Street Configuration

Figure 2. Cluster size distributions of different movement areas follow the exponential function

cluster size gets maximum at the destination and decreases when the distance to the destination increases. Moreover, in the first case (with staying period t_1), nodes are more clustered at the destination since the staying period is longer than the second case (with staying period t_2). We further fit the cluster size distribution to the exponential function in the form of $y = ae^{-cd} + b$ with $d > 0, a > 0, c > 0, b \geq 0$. We see that the cluster size distributions of these two movement areas fit very well with the exponential function. Here, y is the cluster size, d is the distance to the destination, and a, b, c are constants.

In conclusion, when the path to destination is selected such that the shorter distance has a higher probability, the nodes form clusters. Although the cluster size distribution varies, it follows an exponential function with respect to distance to the destination. In the next sections, we leverage this exponential-cluster-size function to design an adaptive content distribution protocol.

IV. System Models

A. Network Model

We focus on clustered mobile P2P networks where the cluster size follows the exponential function. In our network, each mobile node has long-range and short-range communication interfaces. For example, a cell phone may have the cellular communication as the long-range communication and the Bluetooth or wifi communication (i.e., peer-to-peer) as the short-range communication. We assume that the cellular communication has longer delay and lower throughput while the P2P communication has shorter delay and higher throughput. Here, we use the cellular communication to denote the long-range communication and P2P communication to denote the short-range communication.

In our clustered mobile P2P network, we consider the scenario where a mobile user is moving toward a center of a cluster (i.e., the destination) and wants to download a file using her cell phone. She may set the deadline of the file download as the arrival time at the destination. The cell phone (i.e., mobile node) then downloads data from the file server via the cellular communication. At the same time, the node exchanges downloaded data via the P2P communication with other nodes (which are going to the destination as well) to accelerate the download.

Definition of Communication Cluster: In our context the communication cluster is defined recursively as follows:

1) If node n_1 and node n_2 are in the communication range of each other via the P2P channel and are heading to the same destination, then n_1 and n_2 belong to the same communication cluster. In this case, n_1 and n_2 are one-hop neighbors of each other
2) If node n_1 and n_2 belong to the communication cluster C, n_2 and n_3 belong to the communication cluster C, then n_1 and n_3 belong to the communication cluster C. This property means the communication cluster can be expanded to multiple communication hops

In this paper, we use the terms *communication cluster* and *cluster* interchangeably.

B. Data Model and Online Codes

We use Online Codes [14] (or erasure code) in our protocol design. Particularly, the file server divides the original file into B equal-sized *message blocks*. B is a large number so that duplication in block generation can be avoided. For example, a 16MB file can be divided into 2^{13} message blocks and each is 2KB.

Given B message blocks, the server performs the following encoding procedure. The server will first create $Bk\delta$ auxiliary blocks from the B message blocks. To create these auxiliary blocks, each message block will be added by k distinct randomly-chosen auxiliary blocks and each auxiliary block is the sum of $1/\delta$ message blocks on average, where δ is a tunable parameter. The typical value of δ is 0.005 and value of k is 3. These auxiliary blocks are used to facilitate the decoding procedure in the client side. Finally, the server has a set F' of $B' = B + Bk\delta$ blocks.

Upon receiving a request for a block from a mobile node n, the server will create a check block using the B' blocks and send the check block to n. Notice that the check block is the transmitted data unit between mobile nodes and the

Name	Description
F	Number check blocks n must download to decode the original file, $F = (1-\delta)B'$
T_D	Deadline at which n must finish downloading the file
T_C	Current time
T	Current cellular download timer
$g(t)$	The predicted exponential-cluster-size function, $g(t) = ae^{\lambda t} + b$
$G(t_k)$	Set of nodes in the same cluster with n at time t_k, including n. Size of $G(t_k)$ is $g(t_k)$
M	Number of check blocks n is carrying throughout the current time T_C
Δ	The protocol time period. For each period Δ, node n samples cluster size, predicts $g(t)$, and updates T
B_p	Estimated number of check blocks n may obtain from the P2P channel during period $[T_C, T_D]$
B_c	Estimated number of check blocks n has to download from the server during period $[T_C, T_D]$ to meet download deadline

Table I
NOTATIONS USED IN THE PROTOCOL DESIGN

server via the cellular channel, and among nodes in the P2P channel. To create a check block, server picks a random number p (i.e., $p \le B'$) and selects p blocks at random from F'. Then, the server calculates the sum of these p blocks (e.g., server performs XOR operations on p blocks) to obtain the check block q. The server then sends the packet in the form of (x, q) to n, where x is a vector (or meta-data) of all indices of p selected blocks in F'.

To decode the original file, node n collects all received check blocks (x_i, q_i), where the index i denotes the i^{th} check block received by n. Node n then can iteratively recover original blocks from these received check blocks until all B message blocks are decoded. As shown in [14], the B message blocks will be decoded in linear-time as long as n receives at least $(1-\delta)B'$ blocks (either from the server via cellular channel or from other nodes via P2P channel).

V. ADAPTIVE CONTENT DISTRIBUTION PROTOCOL

A. Design Objective and Protocol Overview

The objective of our protocol is to minimize the cellular download and meet the file download deadline for nodes in clustered mobile P2P networks. In order to achieve that, we exploit the clustering behavior to design an adaptive downloading protocol. Particularly, each mobile node n will periodically obtain and keep the cluster sizes in a cluster size list. Then, n predicts the future cluster size function by fitting the cluster size list to an exponential function (i.e., curve fitting) and uses this function to estimate the data availability in P2P channel based on Online Codes [14]. Given the estimation of data availability in P2P channel, n tunes the cellular download timer T to meet the download deadline. The protocol is adaptive since n predicts the cluster size function and tunes the cellular download timer on the move. Table I shows notations used in our protocol design, which will be presented in detail in the following sections.

B. Bootstrapping

In our protocol, we divide time into equal-sized periods Δ. The length of the period Δ depends on node's speed. For example, for the network formed by cellular phones of pedestrians, Δ is longer than that of the network formed by mobile phones of drivers. This is intuitive since when nodes move faster, the cluster size changes faster and thus Δ should be smaller.

When a node n starts moving from its initial location toward the destination, n sets a deadline for the file download and tunes both cellular and P2P channels on. Initially, n has a default cellular download timer T (with $T < \Delta$) and when T expires, n requests data from the content server via the cellular channel. At the same time, n exchanges downloaded data with other peers within n's P2P one-hop communication range. To avoid wrong prediction of cluster size function, n first samples several cluster sizes, one sample per period Δ, and put these cluster sizes into the list of cluster sizes. Then, for each period Δ, n predicts the future cluster size function using the list of cluster sizes, downloads data from cellular channel when T expires, tunes T to meet the file download deadline, and exchanges data via the P2P channel. Next, we present in detail these actions of n.

C. Predicting the Cluster Size Function

As shown in Section III, the cluster size follows an exponential function with respect to the distance d to the destination: $y = ae^{-cd} + b$. Let t denote the time since the node starts moving toward the destination, we see that the distance d is a decreasing function of t (i.e., $d = \frac{e}{t}$, with e is a constant). Let $g(t)$ be the cluster size at time t. Since $y = ae^{-cd} + b$, we have $g(t) = ae^{\lambda t} + b$ with $g(t) > 0, \lambda > 0$. Notice that $g(t)$ is an increasing function with respect to t. The next step is to obtain the coefficients a, b, λ.

For each period Δ, the node n obtains a new cluster size and puts this new cluster size into n's list of cluster sizes. To obtain the cluster size, n broadcasts a membership message, and nodes in n's cluster respond to this membership message. n can obtain the cluster size based on the responsive messages. Then, n fits the list of cluster sizes to an exponential function in the form of $ae^{\lambda t} + b$ to obtain coefficients a, b, λ of $g(t)$. Since nodes choose different paths to the destination, cluster sizes observed by nodes may be different, and thus the coefficients a, b, λ may differ. Node

n will use its predicted exponential-cluster-size function $g(t)$ to estimate the amount of data node n can obtain from the P2P channel (to be shown in Section V-E).

D. Downloading Data From Cellular Channel

Each node n has a cellular download timer T. When the cellular download timer T expires, node n requests new data from the server. Upon receiving the request from n, the server applies the procedure in Section IV-B to create a check block and sends it to n. Since Online Codes is applied at the server side and the number of message blocks B is large, the server will not create duplicate check blocks with high probability [14]. Therefore, any check block returned by the server is useful for n to decode the original file.

E. Tuning the Cellular Download Timer

Given the predicted cluster size function $g(t)$, the next step is to calculate the data availability in P2P channel (i.e., B_p) and tune the cellular download timer T adaptively. In our protocol, when the two nodes first meet, they exchange all new check blocks via the P2P channel. After that, when they stay in one cluster, they only exchange check blocks, which are newly downloaded during the last period Δ. Moreover, for one period Δ, the number of newly downloaded check blocks is not very large. Since the P2P channel has much higher data rate than the cellular channel, we assume that all nodes in one cluster can exchange all newly downloaded check blocks during the last period Δ. For example, the IEEE 802.11 wifi channel currently can support 54Mbps while the cellular channel only support 2Mbps [19]. That means, at the end of the period Δ, all nodes in one cluster share the same set of downloaded check blocks. Then, B_p is estimated as follows.

Since n only knows its current cellular download timer T, n assumes all nodes in n's current cluster have the similar cellular download timer T. Intuitively, nodes in one cluster carry the same set of check blocks and observe similar cluster sizes, so their cellular download timers should be similar. Then, for a future time $t_k > T_C$, the cluster size $g(t_k) = ae^{\lambda t_k} + b$ can be predicted by using the function $g(t)$ of n. Further, the number of check blocks a node in $G(t_k)$ can download for the period $[t_k, t_k + \Delta]$ is $\lfloor \frac{\Delta}{T} \rfloor$. The total number of check blocks all nodes in $G(t_k)$, including n, download for the period $[t_k, t_k + \Delta]$ is $g(t_k)\lfloor \frac{\Delta}{T} \rfloor$. Notice that this calculation holds since the server will not create duplicate check blocks with high probability by using Online Codes [14]. As a result, all new check blocks downloaded by other nodes in $G(t_k)$ are useful for n to decode the original file. The number of period Δ between the current time T_C and the file download deadline T_D is $\sigma = \lfloor \frac{T_D - T_C}{\Delta} \rfloor$. So, we have:

$$B_p = \sum_{i=0}^{\sigma} g(T_C + i\Delta)\lfloor \frac{\Delta}{T} \rfloor \tag{2}$$

As shown in Table I, M is the current number of check blocks that n is carrying throughout T_C. Therefore, $M + B_p$ is the number of check blocks n can potentially obtain by the file download deadline T_D, using the unchanged cellular download timer T for the duration $[T_C, T_D]$. To make the protocol adaptive, the cellular download timer T is updated as follows:

1) If $M + B_p < F$, then $B_c = F - (M + B_p)$. The number of check blocks that n needs to download from the cellular channel for each period Δ is $\lceil \frac{B_c}{\sigma} \rceil$. As a result, the new cellular download timer is $T = \lfloor \frac{\Delta\sigma}{B_c} \rfloor$. If $T > \Delta$, then $T = \Delta$
2) If $M + B_p > F$, then we want to minimize the cellular download. So, we increase the cellular download timer to $T = T + \lfloor \frac{T}{2} \rfloor$. If $T > \Delta$, then $T = \Delta$

F. Exchanging Data via the P2P Communication

A node n uses a bit vector to advertise available check blocks and request for the missing check blocks. For the bit vector , the bit 1 represents the existence of the corresponding check block while the bit 0 means a missing check block. Using the bit vector, n reduces the communication overhead significantly. For example, a file with 10000 check blocks will need 10000 bits, or about 1KB to represent the bit vector.

Receiving the bit vector of its one-hop neighbors in the same cluster, n serves its check blocks, which are missing to its neighbors. To reduce network overhead, nodes only exchange bit vectors and check blocks with their one-hop neighbors. To further reduce network overhead, a node only sends the new bit vector after it obtains new check blocks, either from its neighbors or from the server. The bit vector is exchanged by one-hop neighbors in broadcast communication. Meanwhile, the check blocks are exchanged by one-hop neighbors using the unicast communication to achieve a reliable transmission and higher data rate.

G. Protocol Summary and Discussion

In summary, our protocol divides time into equal-sized period Δ. A mobile node n periodically (for each period Δ) samples the cluster size and adds it into the list of cluster sizes. Then, n fits the list of cluster sizes to an exponential function in the form of $g(t) = ae^{\lambda t} + b$. Then, n uses $g(t)$ function to predict the future cluster sizes for the duration $[T_C, T_D]$ and calculates the number of check blocks (i.e., B_p) n can potentially obtain from the P2P channel during this duration. Next, n calculates the number of check blocks (i.e., B_c) n needs to download from the cellular channel and tunes the cellular download timer T accordingly to meet the file download deadline. When n gets closer to the destination, the cluster size list represents the cluster size distribution of the network better. As a result, the prediction of cluster size function $g(t)$ becomes more accurate and thus the cellular download timer is updated more accurately.

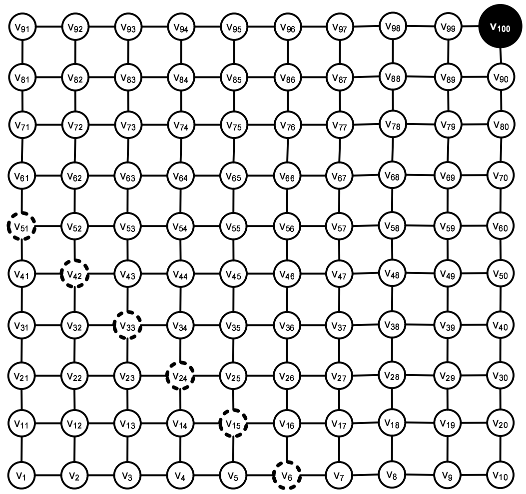


Figure 3. Simulation Area

When n receives over F check blocks, n can decode the original file. Since by using Online Codes, the server does not create duplicate check blocks, all the check blocks that n receives from the P2P channel is useful to decode the file. Notice that since the fitting function of the cluster size is an exponential function, which is more expressive than a linear function, we believe our scheme works even better when the cluster size distribution exhibits a linearly increasing function. Finally, our protocol is not limited to the combination of cellular and wifi/Bluetooth communications. We believe that our protocol is applicable for clustered mobile P2P networks with other communication combinations such as cellular and ZigBee.

VI. EVALUATION

A. Simulation Settings

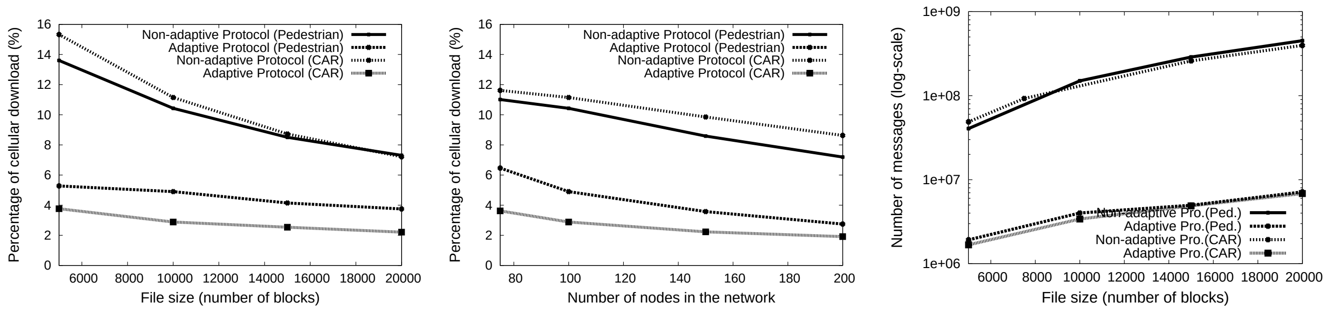
Parameter	**Description**
Number of nodes	[75,100,150,200]
Street seg. length (v_1, v_2)	Pedestrian:100, Car:500 (m)
Destination	v_{100}
Download deadline	1200 (s)
Start locations	$v_6, v_{15}, v_{24}, v_{33}, v_{42}, v_{51}$
File length	[5000,10000,15000,20000](block)
Block size	2KB
Cellular download rate	2 (Mbps) [19]
P2P transmission range	Pedestrian:10, Car:75 (m)
Node speed	Pedestrian:[1,2], Car:[8,10] (m/s)
Δ	Pedestrian:50(s), Car: 40(s)

Table II
SIMULATION SETTINGS

We simulate our protocol in C++. Nodes in our simulation move in a Manhattan street area with 100 intersections as shown in Figure 3. Table II shows the details of simulation settings. For the cellular channel, we use the Proportional Fair Scheduler with the maximum downloading rate 2MBps [19]. Initially, a node n is placed at one of intersections $v_6, v_{15}, v_{24}, v_{33}, v_{42}, v_{51}$. Then, n moves toward the destination v_{100}. To avoid the wrong cluster size sampling, initially n has the initial cluster size 1 and after the first 100(s), n starts sampling the cluster size. n starts to predict the cluster size function at 200(s). At each intersection, n selects the next intersection toward v_{100} by using the probabilistic model in Section III. In our simulation, we have two different types of clustered mobile P2P networks named Pedestrian and Car networks. The former is used to evaluate the performance of our protocol for the pedestrian network with a smaller transmission range and a slower speed. This type of network corresponds to the scenario where football fans are walking to the football stadium to attend a football match as mentioned in Section II, where they download the video of match preview via the cellular channel and exchange data via the Bluetooth channel. Meanwhile, nodes in Car network has a longer transmission range and a faster speed, which corresponds to the scenarios where commuters drive from home to office and download videos of local news via the cellular channel to their cell phones and exchange downloaded data via the wifi channel. For the Car network, we use 75 (m) for the transmission range since the transmission range in practice is much shorter than the theoretical range (i.e., 250 m). The deadline is set to 1200 (s) since a node n is about to arrive at v_{100} with this deadline. We run the experiments 10 times and plot the average.

B. Simulation Results

1) Comparison with Non-adaptive Protocol: Figure 4 compares the performance of our adaptive protocol with a non-adaptive protocol. The non-adaptive protocol uses the P2P channel to exchange downloaded message blocks and at the same time uses the cellular channel to download random message blocks from the content server (See Section IV-B). The non-adaptive protocol does not use bit vectors to advertise and request data. Moreover, the non-adaptive protocol neither uses Online Codes to encode the file nor estimates the data availability in P2P channel. In our simulation, using the adaptive protocol, all nodes finish downloading the file by 1000 (s), before they arrive at the destination v_{100}. Figure 4 shows that the adaptive protocol outperforms the non-adaptive protocol in both Pedestrian and Car networks. Particularly, the adaptive scheme reduces the cellular download from 4% to 10% compared to the non-adaptive scheme, while preserving the file download deadline. In Figure 4(a), when the file size increases, the adaptive protocol works better for Car network since when the network is more dynamic, nodes may meet more peers and have a higher chance to obtain more missing check blocks. Figure 4(b) shows that when the number of nodes in the network increases, both adaptive protocol and non-adaptive protocol perform better since the nodes have more



(a) Variation of file size

(b) Variation of number of nodes

(c) Total message overhead

Figure 4. The adaptive scheme outperforms non-adaptive scheme

peers to exchange data. Again, the adaptive scheme works better for the Car network due to the increases of network dynamics. Figure 4(c) shows that compared to the non-adaptive protocol, the use of bit vectors in block advertisement and request significantly saves the message overhead of our adaptive protocol. Notice that the y-axis of this figure is in log-scale.

In conclusion, our adaptive scheme is robust to node mobility, reduces the cellular download, and reduces message overhead significantly.

2) Fitting Error: Figure 5 shows that when the nodes get closer to the destination (or closer to the download deadline), our protocol can predict $g(t)$ more accurately. In this figure the average normalized error is calculated as follows. For a node n, let Y denote the set of all predicted cluster size functions $g(t)$ of n for the entire simulation. For one $g(t) \in Y$, we calculate the sum of absolute error, E, when comparing $g(t)$ with the cluster size function of n at the file download deadline T_D, $g_{T_D}(t)$ where $g_{T_D}(t) \in Y$. Let $E_{\max}$ denote the maximum value of E for all $g(t) \in Y$. Then, we normalize E (i.e., $E = \frac{E}{E_{\max}}$), hence $0 \leq E \leq 1$ for all $g(t) \in Y$. As a result, for each node n we have a set of normalized sum errors of all $g(t) \in Y$. The plot in Figure 5 is obtained by averaging the normalized sum errors of all the nodes in the simulation. In Figure 5, the average normalized error decreases when time is closer to the file download deadline or node is closer to the destination. That means the cluster size function is predicted more accurately when node is closer to the destination.

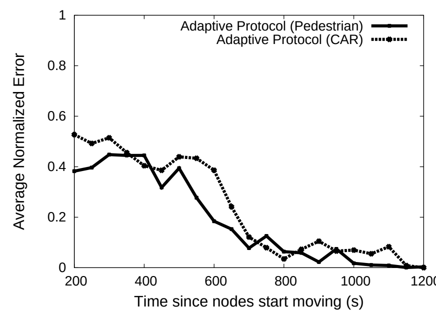


Figure 5. Fitting Error

VII. Related Work

There have been previous projects, which combined cellular and peer-to-peer channels to improve downloading bandwidth of mobile users [1], [3], [4], [5], [6]. These approaches are not adaptive to the network dynamics to reduce the download load on cellular channel, which is one of the most critical issues in mobile networking today, since the deployment of cellular infrastructure is expensive. Our approach is novel since we capture the clustering behavior and predict the P2P data availability on the fly so that we can adapt the download plan efficiently to reduce the load on cellular infrastructure.

Network Coding becomes more and more popular in both wired and wireless P2P content distribution networks [14], [20], [21], [22]. Due to the random nature of check blocks, network coding simplifies the protocol design and improves the protocol performance significantly under network dynamics. In this paper, we use Online Codes at the content server, which provides low redundant check blocks and simplifies the estimation of data availability in the P2P channel.

Current mobility models [7], [8], [9], [23] do not capture the socioeconomic characteristics in human movement [10], [11]. Meanwhile, previous studies have shown that human movement depends on various socioeconomic factors such as weather [11], religious affiliation [24], and social characteristics [10]. Recent measurement studies in [12], [18] show that people do not always choose the shortest path to the destinations due to their personal characteristics [18]. The study [12] also shows that in a city the movement of people form clusters with a much higher density than other areas. This motivates our study of the probabilistic model and clustering behavior in this paper.

VIII. Conclusion

In this paper, we present realistic classes of scenarios where the movement of people forms clustered mobile P2P networks. In these networks, the density of nodes increase significantly at the cluster centers and decreases gradually at farther distance. To understand the clustering behavior, we propose a probabilistic model to represent the path selection and use the Mobius Tool to study the cluster behavior of different street configurations. We find that the cluster size distribution follows an exponential function. Then, we present a novel adaptive protocol that blends cellular and P2P communications of the mobile devices and leverage the exponential-cluster-size function to improve content distribution services of clustered mobile P2P networks. With our protocol, mobile nodes periodically sample the cluster size and predict the future cluster size using the exponential function. Moreover, mobile nodes apply the exponential-cluster-size function to calculate the available data in the P2P channel using Online Codes and tune the cellular download timer adaptively to meet the file download deadline. The simulation results show that our adaptive protocol considerably outperforms non-adaptive protocol and adapts well to network dynamics.

REFERENCES

[1] H. Luo, R. Ramjee, P. Sinha, L. Li, and S. Lu, "The design and evaluation of unified cellular and ad hoc networks," *IEEE Transactions on Mobile Computing*, 2007.

[2] S. G. Aly and A. A. Elnahas, "Sustained service lookup in areas of sudden dense population," *Wireless Communications and Mobile Computing*, vol. 8, pp. 61–74, 2008.

[3] K. S. and M. M., "Efficient mobile access to internet data via a wireless peer-to-peer network," in *Pervasive Computing*, 2004.

[4] S. Jung, U. Lee, A. Chang, D.-K. Cho, and M. Gerla, "Bluetorrent: Cooperative content sharing for bluetooth users," *Pervasive and Mobile Computing*, vol. 3, pp. 609–634, 2007.

[5] H. Yoon, J. Kim, F. Tany, and R. Hsieh, "On-demand video streaming in mobile opportunistic networks," in *Proceedings of Percom*, 2008.

[6] H. Hanano, Y. Murata, N. Shibata, K. Yasumoto, and M. Ito, "Video ads dissemination through wifi-cellular hybrid networks," in *Proceedings of Percom*, 2009.

[7] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *Proceedings of Infocomm*, 2005.

[8] X. Hong, M. Gerla, G. Pei, and C. chuan Chiang, "A group mobility model for ad hoc wireless networks," in *Proceedings of ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, 1999, pp. 53–60.

[9] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek, and M. Degermark, "Scenario-based performance analysis of routing protocols for mobile ad-hoc networks," in *Proceedings of MobiCom*. ACM Press, 1999, pp. 195–206.

[10] A. Papatheodorou, "Why people travel to different places," *Annals of Tourism Research*, vol. 28, no. 1, pp. 164–179, 2001.

[11] O. Jrv, A. Aasa, R. Ahas, and E. Saluveer, *Weather dependence of tourist's spatial behavior and destination choices: case study with passive mobile positioning data in Estonia*, 2007.

[12] *Location Based Services and TeleCartography II (Mapping Pedestrian Movement: Using Tracking Technologies in Koblenz)*. Springer Berlin Heidelberg, 2009, pp. 95–118.

[13] D. D. Deavours, G. Clark, T. Courtney, D. Daly, S. Derisavi, J. M. Doyle, W. H. Sanders, and P. G. Webster, "The mobius framework and its implementation." *IEEE Transactions on Software Engineering*, vol. 28, no. 10, pp. 956–969, October 2007.

[14] P. Maymounkov and D. Mazieres, "Rateless codes and big downloads," in *Proceedings of IPTPS03*, 2003.

[15] M. Hollick, T. Krop, H.-P. Huth, and R. Steinmetz, "Modeling mobility and workload for wireless metropolitan area networks," *Computer Communications*, vol. 27, pp. 751–761, May 2004.

[16] F. Ekman, A. Kernen, J. Karvo, and J. Ott, "Working day movement model," in *Proceedings of MobilityModels*, 2008.

[17] M. Pirkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "On clustering phenomenon in mobile partitioned networks," in *Proceedings of 1st ACM SIGMOBILE workshop on Mobility models*, 2008.

[18] *Location Based Services and TeleCartography II (Ways of Walking - Developing a Pedestrian Typology for Personalised Mobile Information)*. Springer Berlin Heidelberg, 2009, pp. 79–94.

[19] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for normadic users," *IEEE Communications Magazine*, 2000.

[20] C. Gkantsidis and P. Rodriguez, "Network coding for large scale content distribution," in *Proceedings of Infocomm*, 2005.

[21] M. Wang and B. Li, "R2: Random push with random network coding in live peer-to-peer streaming," *IEEE Journal on Selected Areas in Communications, Special Issue on Advances in Peer-to-Peer Streaming Systems*, vol. 25, December 2007.

[22] U. Lee, J.-S. Park, S.-H. Lee, W. W. Ro, G. Pau, and M. Gerla, "Efficient peer-to-peer file sharing using network coding in manet," *Journal of Communications and Networks (JCN), Special Issue on Network Coding*, vol. 10, pp. 422–429, December 2008.

[23] M. Musolesi and C. Mascolo, "A community based mobility model for ad hoc network research," in *Proceedings of International Symposium on Mobile Ad Hoc Networking and Computing*, 2006.

[24] Y. Mansfeld and K. Ya"acoub, "Patterns of tourist destination-choice and travel behaviour among members of the urban and rural arab community of israel: A comparative study of haifa and ibilin," *GeoJournal*, vol. 35, no. 4, pp. 459–470, 2004.